



Pearson



Arizona's Instrument to Measure Standards Science Mode Comparability Report

June 2018

Prepared by Pearson in association with the Arizona Department of Education Assessment Section.

Arizona's Instrument to Measure Standards Science Mode Comparability Report

Hirotaaka Fukuhara and Lietta Scott

June 2018

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Fukuhara, H., & Scott, L. (2018). *Arizona's Instrument to Measure Standards Science mode comparability report*. Phoenix, AZ: Arizona Department of Education, Assessment Section.

Executive Summary

A mode comparability study was performed on the Spring 2018 Arizona's Instrument to Measurement Standards (AIMS) Science assessment as this was its first operational online administration. Item parameter drift due to the mode change was investigated on one of three administered online forms using sample sizes in excess of 28,000. Arizona's standard equating process, displacement by Winsteps (Linacre, 2015), revealed that the vast majority of the items on the tests were stable across modes. The items on the form on which the mode comparability study was performed then became the anchors for equating the other two forms administered, again using Arizona's standard equating method. Impact data comparing the 2018 forms to previous years indicated that there was no shift on student performance due to the mode change.

Introduction

Beginning Spring 2018, the Arizona Department of Education (ADE) moved its Science assessment, the Arizona's Instrument to Measure Standards (AIMS), from a paper-and-pencil to an online administration, with paper accommodated forms available as needed for students with disabilities that prevented them from accessing online content. The AIMS Science is administered in Grades 4, 8, and High School.

ADE also decided that three forms will be spiraled at a student level for the online administration for each of the next three years to increase test security; whereas, for the associated accommodated paper tests (paper, large print, and Braille) only one form will be available. We expected that the vast majority of students will take the tests online and only a handful of students would take them in an accommodated paper format.

One of the online forms in each grade level administered in Spring 2018 contained the exact 4-option multiple-choice items in the exact location of where they were administered during the Spring 2017 administration of the paper-and pencil tests. This online 2017 form was identified as Form C. The two other forms administer at each grade level also contained only 4-option multiple-choice items with a high proportion of items that were identical to those in Form C and were in the same location. These other two forms were designated as forms A and B.

Mode Comparability Study

Since it was the expectation that items administered in the paper-and-pencil format in the past would behave similarly in an online format, a mode comparability study was performed via Arizona's standard linking method, in Spring 2018. There were a couple of reasons this method was chosen. First, a comparability study for AIMS Science has been conducted once in 2007 (Arizona Department of Education, 2007). This study, which was based on a stand-alone field-test, it did not find significant mode effects. Second, AIMS Science items are all multiple-choice items so that they are not expected to be presented differently between a paper-and-pencil form and an online form. Third, the vast majority of students were expected to take the online assessment in Spring 2018. The few students who may take a paper-and-pencil form (Braille, Large Print, or Paper) may not be a representative of population as these students would have the need for this specific accommodation spelled out in their Individual Education Plan (IEP) or 504 Plan.

Forms A, B, and C were constructed to purposefully have many items overlapping among the forms. These items were available as possible anchors to equate the forms. One issue with this plan was that some of the common items across the forms are at different positions. To facilitate their use as anchors, ideally, some of them would be in the same positions for the online forms. Item position for the common items in Forms A, B, and C (Spring 2017) is presented by grade in the Appendix A. The items that are in the same positions across the three forms are bold in the appendix.

One objective of equating online forms to the original scale, which is based on a paper-and-pencil administration, is to make an adjustment for the mode change so that scores from the online forms will be reported on the existing AIMS Science scale. A two-step procedure was implemented to achieve the objective. A graphical presentation of steps for equating online forms is presented in Figure 1.

Step1: Mode Adjustment

The first step was to conduct a mode adjustment study between a paper and an online form in Spring 2018. We made the adjustment through a linking method on Form C, an intact form from Spring 2017, which was administered in a paper-and-pencil format in Spring 2017 and delivered online in Spring 2018. For this exploration of mode effect, item drift was examined using displacement within Winsteps (Linacre, 2015) where item difficulty for AIMS dichotomously scored items is modeled using the Rasch model (Rasch, 1960). AIMS' standard, fixed-anchor, non-equivalent groups anchor item (NEAT) method, was used except rather than a subset of items used as anchors, all operational items were used and examined for displacement. In the literature, a displacement (change in difficulty from the fixed value to the new value, if the item were freely estimated) of greater than 0.5 logits in magnitude is of concern (Linacre, 2018). Arizona, however, flags any item with a value of displacement greater than 0.3 in magnitude, so that only anchor items that have estimated difficulty values within approximately 1/3 logit of their fixed value are maintained as anchors. Those that do not meet this threshold, are released from the anchor set and freely estimated in an iterative process releasing the item with the largest flagged displacement and re-equating the test until no more anchors are flagged for displacement.

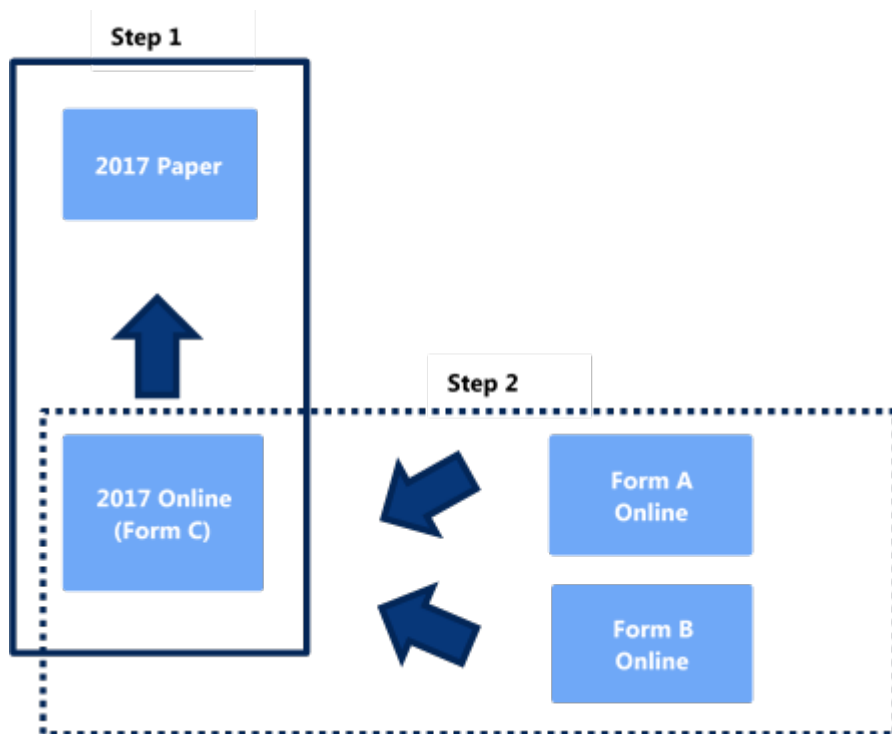


Figure 1. Graphical Presentation of Steps on Online Form Equating

Step 2: Equate Online Forms

Once Form C was linked to the existing AIMS Science scale, the next step was to equate Forms A and B, separately, to Form C to put them also on the original AIMS Science scale. In other words, two separate equating analyses were run, one to equate Form A to the newly equated Form C and the other to equate the Form B to the newly equated Form C, rather than equating them simultaneously. This separate calibration, again using the NEAT method, was necessary because some of common items across forms were placed at very different positions. These “displaced” items were not treated as the ‘same’ items as their item parameter estimates were expected to be affected by their major shift in placement (Miller & Fitzpatrick, 2009). All common items that were at the same location as where they were on Form C were used as anchors across forms.

Results

A mode comparability study, via linking, revealed that the vast majority of items were stable by mode. For Grade 4, only 2 out of 54 items showed displacement with an absolute value larger than 0.3. One item became relatively easier than its 2017 value (Item 3: displacement of -0.4680) while 1 item became relatively more difficult than its 2017 value (Item 2: displacement of 0.3166). For Grade 8, 2 out of 58 items also showed displacement outside of the criteria. Both of them became relatively more difficult than their 2017 values (Item 44: displacement of 0.4704 and Item 16: displacement of 0.3126). In High School, there were no items flagged for displacement. A scatterplot of *p*-values for Spring 2018 Form C vs. the Spring 2017 administration is presented in Appendix B.

Once all of the item parameter estimates from Form C were on each grade level’s AIMS base scale (the purpose of Step 1), common items at the same location among all online forms for each grade level were used as anchors to equate Forms A and B to their respective base scales. In this step, the only item that was flagged for displacement appeared in the Grade 4 Form B. This item (Item 43: displacement of -0.3276) was dropped from the anchor set on that form.

After Step 1 and 2 were complete, an impact analysis was conducted on both test characteristics and student performance for the 2018 online forms by comparing the results against a historical trend as a reasonableness check. The historical trend was summarized by grade level in Appendix C. The analysis for High School was further broken down by cohort since students in Grade 9 were allowed to take the “Grade 10” test if they are enrolled in a life science course that is aligned to Strands 1-4 of the Arizona Academic Content Standard for Science at the high school level (ADE, 2018). Note that there was only 1 core form for each grade level in 2015 through 2017. In terms of test characteristics, the average *p*-value and Rasch difficulty values for the 2018 online forms were comparable to the previous years. Consequently, raw score cuts for the 2018 online forms were very close to the previous years. Similarly, student performance on the 2018 forms were comparable to the previous years for all grades with respect to the average scale score and Performance Level distribution.

Conclusions

The mode comparability study from Step 1 revealed that most of the operational items were stable across modes. The few items that did show item drift due to the mode change, were freely calibrated so that their item parameter estimates were updated to the base scale. Impact data showed that student performance in Spring 2018 was similar to Spring 2017 in terms of average scale score and Performance Level distribution. These results suggested that there was no shift in student performance due to the transition to online testing.

References

- Arizona Department of Education. (2018). *Spring 2018 Test Coordinator's Manual, AIMS Science Grades 4, 8, and High School*. Phoenix, AZ: Arizona Department of Education Assessment Section. Available from www.azed.gov.
- Linacre, J. M. (2015) Winsteps® (Version 3.90.0) [Computer Software]. Beaverton, OR: Winsteps.com. Retrieved January 1, 2015. Available from <http://www.winsteps.com/>.
- Linacre, J. M. (2018). A user's guide to WINSTEPS® MINISTEP Rasch-model computer programs 4.1.0. Beaverton, OR: Winsteps.
- Miller, G.E. & Fitzpatrick, S. J. (2009). Expected equating error resulting from incorrect handling of item parameter drift among the common items. *Educational and Psychological Measurement*, 69(3), 357-368.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedagogiske Institut.

Appendix A. Item Sequence for Common Items on Forms A, B, and C

Grade 4

AZID	Form A	Form B	Form C (Spring 2017)
44134452	001	001	001
44134453	002	002	002
44104271	003	003	003
44104269	004	004	004
44104270	005	005	005
3514438	006	006	006
3514583	007	007	007
44134455	008	008	008
44134461	009	009	009
3514437	013	013	013
44134451	018	018	018
44144018	019	017	017
3514430	020	019	019
3514569	021	020	020
3514484	022	021	021
3514504	023	022	023
44134448	024	024	024
44114450	025	023	022
44114437	028	028	028
44114434	029	029	029
44114327	030	030	030
44144026	031	031	027
44124314	032	032	032
3514603	033	033	033
3514485	034	034	034
44134487	035	035	035
44134478	036	036	036
44134449	037	037	037
3514444	039	039	039
44104309	040	040	040
3514452	042	042	042
3514435	043	043	043
44144009	044	044	047
44124273	045	026	031
3514477	046	046	045
44114444	047	047	014
44114443	048	048	015
44114447	049	049	016
44104068	050	050	050
44104069	051	051	051
44104077	052	052	052
44104334	053	053	053
44134473	054	054	054

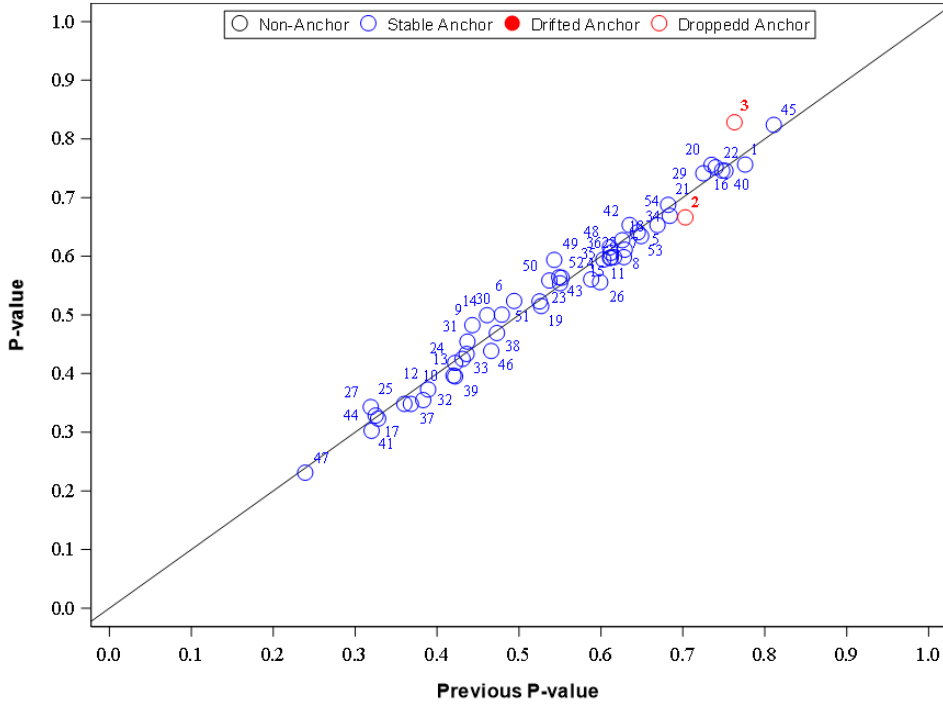
Grade 8

AZID	Form A	Form B	Form C (Spring 2017)
44108415	001	001	001
44108391	002	002	002
3518780	007	007	007
3518615	008	008	008
3518786	009	009	009
3518788	010	010	010
44138121	012	012	012
44128767	013	011	011
44138394	014	013	013
44138005	015	014	014
3518794	016	016	016
3518776	017	017	017
44138415	018	015	015
3518785	019	018	019
44128354	020	020	020
44138129	023	023	023
44138127	024	024	024
3518701	025	022	022
44138401	026	026	026
44148057	027	025	025
44138122	028	028	028
3518724	029	029	029
44138385	030	030	030
44138390	031	031	031
44138405	032	032	032
3518630	034	033	033
3518628	035	034	034
44148039	036	035	037
44148377	037	027	027
3518640	038	038	038
3518613	039	036	035
3518703	040	040	040
3518706	041	039	039
3518705	042	041	041
3518768	043	043	043
44108984	044	044	044
44108410	045	045	045
44108988	046	046	046
44108369	047	053	056
3518801	048	048	048
3518722	049	051	049
3518674	050	005	005
3518675	051	047	053
3518791	052	052	052
44108349	054	054	054
44148024	055	050	051
44148022	056	055	055
44128311	057	057	057
44128314	058	058	058

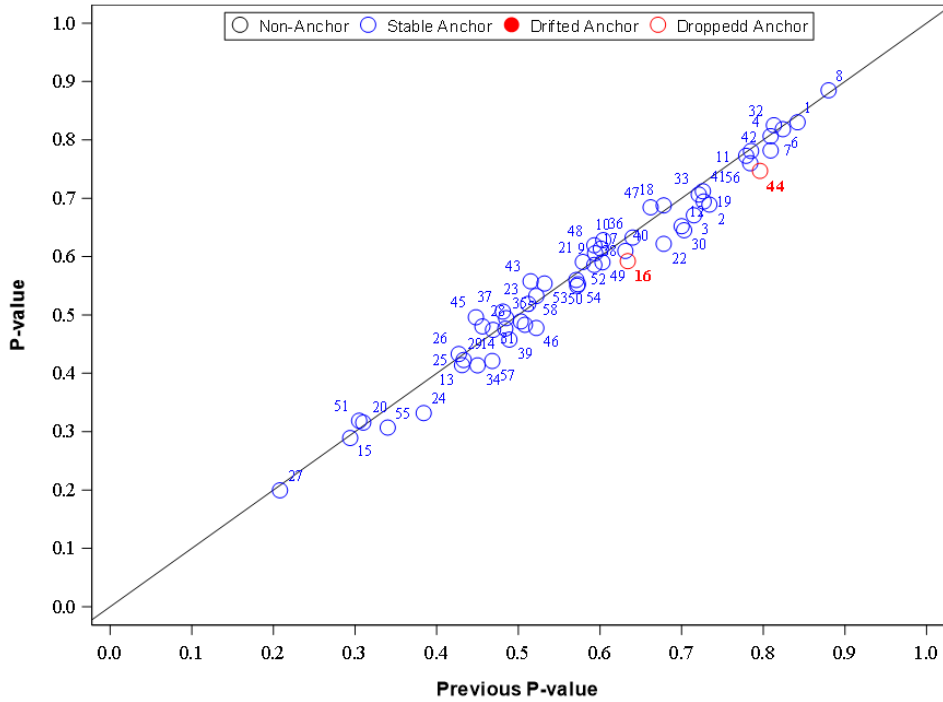
High School

AZID	Form A	Form B	Form C (Spring 2017)
44130258	001	001	001
44100601	002	002	002
44100610	004	004	004
3510037	005	005	005
3510193	008	008	008
44100819	009	009	009
44100622	010	010	010
44100818	011	011	011
44140010	012	012	012
3510132	013	013	013
3510134	014	014	014
3510136	015	015	015
44140011	016	016	016
3510024	017	017	018
44140039	018	018	020
44120381	019	025	025
44140012	020	019	017
3510092	022	022	022
3510093	023	023	023
3510090	024	021	021
44120048	025	024	024
3510074	026	029	026
44130291	027	026	029
44130293	028	027	030
44130285	029	028	031
3510075	031	031	027
44130272	032	032	032
44130284	033	033	033
3510200	034	034	028
44140028	035	035	060
3510191	036	036	036
3510111	037	042	041
44140041	038	041	035
3510166	039	040	040
44100565	040	043	042
44130248	041	037	037
44130245	042	038	038
44130247	043	039	039
44120359	044	044	044
44120288	045	045	045
3510083	046	047	046
3510141	047	046	047
44140045	048	048	049
3510031	050	050	050
44130287	051	052	065
44130144	052	056	048
44100642	053	051	063
3510174	054	054	055
44130206	056	059	059
44130213	057	057	057
44130205	058	058	058
3510045	059	063	053
3510115	061	061	061
44100583	062	062	062
3510140	064	049	054
44130283	065	065	064

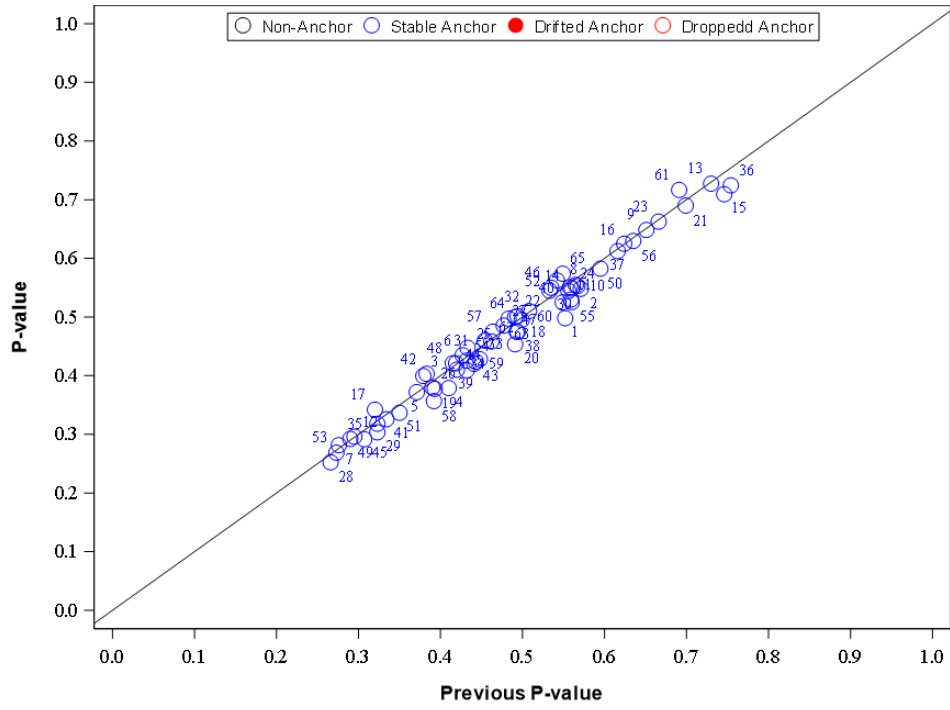
Appendix B. Scatterplot of P-value from Spring 2018 Form C vs Spring 2017



(a) Grade 4



(b) Grade 8



(c) High School

Appendix C. Historical Trend in Test Characteristics and Student Performance

Table C.1.a. Historical Trend in Test Characteristics for Grade 4

Year	Form	Average P-value	Average Rasch	Raw Score Cuts		
				Approaches	Meets	Exceeds
2018	*Form C (2017)	0.55	0.6868	18	27	38
	Form B	0.56	0.6103	19	28	39
	Form A	0.55	0.6247	19	28	39
2017	Paper	0.55	0.6892	18	27	38
2016	Paper	0.56	0.6187	19	28	39
2015	Paper	0.55	0.6407	18	28	38

* This form was studied for mode comparability.

Table C.1.b. Historical Trend in Student Performance for Grade 4

Year	Form	N	Average Scale Score	SD Scale Score	Percent at Performance Level			
					Falls Far Below	Approaches	Meets	Exceeds
2018	*Form C (2017)	29443	514.50	45.61	12	28	36	24
	Form B	29423	513.94	47.17	14	27	34	24
	Form A	29383	512.79	45.82	14	28	36	22
2017	Paper	87350	514.96	46.76	13	27	35	25
2016	Paper	85675	514.49	47.56	15	25	34	25
2015	Paper	83905	513.86	46.53	13	29	32	26

* This form was studied for mode comparability.

Table C.2.a. Historical Trend in Test Characteristics for Grade 8

Year	Form	Average P-value	Average Rasch	Raw Score Cuts		
				Approaches	Meets	Exceeds
2018	*Form C (2017)	0.58	0.4050	25	32	40
	Form B	0.57	0.4595	24	31	39
	Form A	0.56	0.4798	24	31	39
2017	Paper	0.58	0.3913	25	32	40
2016	Paper	0.57	0.4299	24	31	39
2015	Paper	0.57	0.4494	24	31	39

* This form was studied for mode comparability.

Table C.2.b. Historical Trend in Student Performance for Grade 8

Year	Form	N	Average Scale Score	SD Scale Score	Percent at Performance Level			
					Falls Far Below	Approaches	Meets	Exceeds
2018	*Form C (2017)	28310	512.13	50.27	25	20	23	32
	Form B	28154	512.25	49.50	24	20	23	34
	Form A	28230	512.19	48.41	24	20	23	33
2017	Paper	83398	514.00	50.85	23	19	24	34
2016	Paper	82258	512.64	48.83	22	19	25	35
2015	Paper	82042	513.06	48.08	22	20	24	34

* This form was studied for mode comparability.

Table C.3.a. Historical Trend in Test Characteristics for High School

Year	Form	Average P-value	Average Rasch	Raw Score Cuts		
				Approaches	Meets	Exceeds
2018	*Form C (2017)	0.48	0.5651	28	35	45
	Form B	0.49	0.5031	28	36	45
	Form A	0.49	0.4850	29	36	46
2017	Paper	0.48	0.5651	28	35	45
2016	Paper	0.49	0.5240	28	35	45
2015	Paper	0.50	0.5135	28	35	45

* This form was studied for mode comparability.

Table C.3.b. Historical Trend in Student Performance for High School

	Year	Form	N	Average Scale Score	SD Scale Score	Percent at Performance Level			
						Falls Far Below	Approaches	Meets	Exceeds
10 th Grade	2018	*Form C (2017)	15362	479.60	47.19	54	17	17	12
		Form B	15186	479.01	46.34	51	21	15	13
		Form A	15526	478.61	45.83	53	18	18	11
	2017	Paper	44191	483.30	46.99	50	18	18	13
	2016	Paper	46242	482.21	44.16	48	20	19	12
	2015	Paper	50767	484.29	44.73	45	20	21	14
9 th Grade	2018	*Form C (2017)	12940	499.43	50.54	36	19	23	22
		Form B	12809	498.46	49.30	34	22	21	23
		Form A	12927	499.41	48.69	35	19	25	22
	2017	Paper	36104	498.93	49.68	36	19	23	22
	2016	Paper	33782	499.41	48.36	33	20	25	22
	2015	Paper	28869	504.20	49.28	29	19	26	26

* This form was studied for mode comparability.